

АКТУАЛЬНОСТЬ И МЕТОДИЧЕСКИЕ АСПЕКТЫ ОБУЧЕНИЯ СЕМАНТИЧЕСКИМ СЕТЯМ В ПЕДАГОГИЧЕСКОМ ВУЗЕ

Ефимов А.А., Рожина И.В.

*Уральский государственный педагогический университет
Екатеринбург, Россия.*

Аннотация

Авторы статьи пытаются обосновать необходимость обучения студентов созданию семантических сервисов. Он считает, что семантическая сеть перспективная молодая технология снабжения всех данных метаданными, будет востребована в будущем и породит большой спрос на специалистов, умеющих не только эффективно использовать. Но и конструировать семантические сервисы. На примере XML и XHTML – основных языков разметки в рамках семантических сервисов он раскрывает простоту семантической сети.

TOPICALITY AND METHODOICAL ASPECTS OF TEACHING SEMANTIC WEB IN PEDAGOGICAL UNIVERSITY

A. Efimov, I. Rogina.

*Ural State Pedagogical University
Yekaterinburg, Russia.*

Abstract

The author of the article tries to prove the essentiality of teaching semantic web students of the Information Department of the Ural State Pedagogical University. He states that though the semantic web is a relatively new the technology of providing all data with metadata is topical and essential and in future there will be a great demand for specialists who can not only use the semantic services, but can create them. By the example of XML and XHTML, basic markup languages of the semantic web, he reveals the simplicity of the semantic web concept.

Эволюция сетей: от компьютерных и информационных к семантическим

Сеть «Интернет» прошла два этапа своей эволюции. Первым этапом, уходящим в 70-ые годы, был собственно Интернет — выросшая из военного проекта глобальная компьютерная сеть, узлы которой обменивались данными, не требовавшими обработки в реальном времени. В начале 90-ых родилась WORLD WIDE WEB — Всемирная паутина, которая предоставила унифицированный способ доступа к различным типам данных. В результате этого, вместилищем информации стали не файловые архивы, а WEB-сайты, работа online сделалась проще, наглядней и интересней, а сама «сеть» превратилась из глобальной

компьютерной в глобальную информационную. Но и WEB не стал последним этапом: на очереди следующая эволюционная ступень, именуемая Semantic WEB — семантическая сеть.

Семантическая сеть была задумана консорциумом W3 в середине 90-х годов. В 2001 году в журнале Scientific American появилась статья Тима Бернерса-Ли, Джеймса Хендлера и Оры Лассила под названием «The Semantic WEB». Тим Бернерс-Ли — это отец WWW. Он не только первым предложил концепцию всемирной паутины, но и воплотил её в жизнь. К словам такого человека просто нельзя не прислушиваться.

Стиль описания похож на пророчества манифеста «Some Computer Science Issues in Ubiquitous Computing» [3], которые, кстати, сбываются. Например, о том, что главной проблемой является не повсеместная компьютеризация жизни, а конфиденциальность личной жизни в таких условиях.

Под семантической сетью понимается структура, наполненная информацией о смысловом содержании, понятной компьютерам. Современные WEB-страницы созданы для человека, компьютерным программам не под силу проанализировать их содержимое. Главной задачей семантической сети станет добавление специальных данных, которые позволят компьютерным программам понимать WEB-страницы. Такое добавление выведет Интернет на новый виток эволюции. Браузеры научатся «понимать», о чём говорится на WEB-страничках, которые они открывают. Это сразу же на порядок повысит эффективность работы с WEB-ресурсами: компьютеры сами, без помощи человека смогут находить нужную нам информацию, самостоятельно отсеивать сор, похожий словарно, но отличный по смыслу, будут в состоянии автоматически выполнять самые различные поручения по работе с WEB-контентом, вроде заказа мест в ресторане или подбора рефератов. Человеческий интеллект, который сегодня является необходимым инструментом при работе с Сетью, станет машинам не нужен.

На самом деле всё не так фантастично как кажется. Для того чтобы выделить имя или e-mail адрес средствами HTML были созданы тэги <NAME> и <ADDRESS>, но на их стандартизацию ушли годы, а многие из таких тэгов, содержащих «информацию об информации» так и не стали стандартом. Решение множества таких вот «маленьких» проблем и есть Semantic WEB.

На данный момент специалисты, понимающие принципы работы семантической сети, не сомневаются в преимуществах этого решения или в его перспективах. Но, сейчас не существует завершённой технологии разработки сервисов в рамках концепции семантической сети. Перед разработчиками встаёт вопрос, изобретать что-то самостоятельно, или дождаться появления завершённых инструментов, а перед педагогами, обучать этому или нет, но чаще — как именно обучать.

Необходимость изучения семантических сетей

Семантическая сеть — технология молодая, не имеющая достаточного количества примеров успешной реализации, но существующие примеры заслуживают внимания, например, проекты Google Apps или Wiki.

Технология снабжения всех данных метаданными, упрощающими машинный анализ, безусловно, актуальна и востребована. Сейчас семантические сервисы предлагаются в основном крупными компаниями, более того, структура сервиса – интеллектуальная собственность этой компании. В будущем, когда технология шагнёт в массы, рынку понадобятся специалисты не только способные пользоваться семантическими сервисами, но и умеющие их создавать.

Рынок поставщиков программного обеспечения, откликается на перспективные потребности так же быстро, как и рынок поставщиков сервисов. Все разработчики специализированного программного обеспечения, предоставили инструменты разработки на языках, рекомендованных консорциумом W3 для реализации семантических сервисов. Это и технология Silverlight с собственной средой разработки от Microsoft, и технология AJAX, используемая и поддерживаемая сообществом, и многие другие технологии и среды разработки от компаний уровня Adobe и IBM. То есть, семантическая сеть на столько интересна потенциальным пользователям, и количество разработчиков, готовых к реализации сервисов так велико, что крупные поставщики программных продуктов готовы предоставить нужные инструменты для разработки.

Из сказанного можно сделать вывод, о необходимости изучать семантические технологии.

Использование логики силлогизмов при описании данных метаданными

На данный момент, самый распространённый вариант использования идей семантической сети – это такой способ описания данных метаданными, который позволил бы машине использовать логику силлогизмов. Он относительно прост в реализации, поскольку силлогизм знаком человеку со времён Аристотеля. Причем если следовать минимальным требованиям, таким как равноценность понятий, отсутствие софизмов и пр., то и выгоды от использования очевидны, и техническая реализация достаточно проста. А раз для реализации используется язык eXtensible Markup Language (XML), то любой семантический сервис сможет правильно проанализировать документ, ведь язык формализован, и структура документа полностью описана.

Хотя силлогизмы и окружают нас со времен античных философов, они достигли своего апофеоза в 19-ом столетии в работах Чарльза Доджсона (более известного как Льюис Кэрролл). Доджсон написал две книги о силлогизмах и методах их представления в графической форме, и его силлогизмы часто принимали форму сорита, при котором вывод из одной пары связанных утверждений становился новым утверждением, пригодным для связи с другими высказываниями [2].

Один из соритов Доджсона говорит:

Англичане — мужественный народ.
Мужественный народ свободен.

Свободный народ счастлив.
Следовательно, англичане счастливы

Это наводит на мысль о том, что не должным образом сформированный документ может ввести компьютер в заблуждение, результатом которого станет неадекватное действие, будь то просто неверный ответ поисковой машины или хищение пользовательских данных.

То есть, несмотря на привлекательную простоту, силлогизмы не слишком хорошо работают в реальном мире, потому что большинство данных, которыми мы пользуемся, не поддаются таким простым способам обработки. Как результат, семантическая сеть в нынешнем её состоянии не сможет быть однозначно полезной.

Использование дедуктивного способа вывода знаний

Авторы концепции семантической сети существенно переоценивают значение дедуктивного способа вывода знаний. Величайшим популяризатором этой ошибки был Артур Конан Дойл, чьи истории о Шерлоке Холмсе нанесли большой урон нашему пониманию человеческого разума. Дойл убедил поколения читателей, в том, что умные люди, приходят к неизбежно верным заключениям посредством связывания последовательных фактов. Как Холмс лихо утверждал, «когда вы устранили невозможное, то, оставшееся, и будет истиной, как бы неправдоподобно это не выглядело». В реальном мире люди оперируют неполной, неубедительной или чувствительной к контексту информацией. Поэтому, когда приходится принимать решение, основываясь на такой информации, человеку приходится строить догадки, экстраполировать, обращаться к интуиции, поступать так, как поступал ранее в аналогичной ситуации, делать то, что сделали бы друзья.

И как следствие, почти ни одно утверждение, из тех, что делает человек, даже самое очевидное, не является истиной в том смысле, как этого требует семантическая сеть. Дрю МакДермотт в его блестящей «Критике чистого разума» [1] взял под сомнение, что возможно создать искусственный интеллект, построив достаточно детализированную дедуктивную машину. Он заключил, что этот подход фатально недоработан, и заметил, что: - «Это, должно быть случай, когда нам просто хочется, чтобы значительная часть логического вывода была дедукцией, в противном случае это просто не будет соответствовать количеству теорем, дедуктивно выведенных из заданного набора аксиом». Хотя «Критика чистого разума» предворяла не только семантическую сеть, но и «обычный» WWW, критика все еще актуальна.

Из вышесказанного трудно сделать однозначные выводы относительно того не только как обучать, но и собственно, чему. Инструментов великое множество. Единообразие языков разработки только внешнее, концепции нет, практическая польза неоднозначна. Однако, поскольку такие крупные корпорации как IBM и Microsoft проявляют интерес к SilverLight и NetBeans, то и концепция будет сформирована, и варианты практической реализации проработаны, и

положительные потребительские качества станут очевидными.

Например, концепция таксономии - самая естественная классификация. Все остальные можно автоматически вывести из нее, исходя из свойств и отношений.

Дескриптивные логики, исходят из того, что все возможно, пока не доказано обратное. Это не закрытый мир и силлогизмы нужны, в первую очередь не для моделирования мышления, а для дополнительной связи отношений и онтологий, то есть, выстраивания сети семантики для получения адекватных запросу результатов. Для борьбы с субъективностью информации давно существуют свои технологии, например, модальная логика, нечеткая логика, факторы уверенности. Представления о мире отличаются у всех, но это не мешает успешному взаимодействию на том или ином уровне. А на каком, зависит от целей и общности интересов. Так пусть будет больше онтологий, больше разных проекций. Безусловно, на данный момент, технология семантической сети - это не для всех.

Из вышесказанного становится ясно, что педагог, должен уметь видеть проблемные моменты новых технологий и при этом уметь объективно оценить их перспективность. Ведь проблемы рано или поздно разрешатся, а актуальность останется. Кроме этого, педагог должен уметь вычленять нечто общее, основополагающее в новой перспективной технологии, нечто такое, что не претерпит кардинальных изменений в процессе становления технологии и не потеряет практической значимости.

Язык XML – базовый язык разработки WEB приложений

В концепции семантической сети этой основой являются в равной части логика и язык XML. Язык XML – базовый язык разработки WEB приложений настоящего и будущего.

К достоинствам языка можно отнести:

- понятность формата и человеку и компьютеру;
- возможность описания основных структур данных (записей, списков, деревьев);
- самодокументируемость формата, который описывает структуру и имена полей также как и значения полей;
- синтаксис и требования к анализу, которые позволяют языку оставаться простым, эффективным и непротиворечивым.
- иерархическая структура, которая подходит для описания практически любых типов документов.
- независимость от платформы.
- отсутствие требований на расположение символов на строке.

Все сопутствующие языки, для реализации семантических сервисов, такие как язык описания ссылок, язык таблиц стилей, язык шаблона документов, язык для описания баз данных, и многие другие, имеют XML синтаксис, что приводит всю разработку к одному общему знаменателю и упрощает труд разработчиков.

Заключение

Сети семантической, то есть такой, в которой машина сможет анализировать информацию, быть. На данный момент речь не идёт об осмыслении информации, самосознании машины или другой фантастике. Речь идёт о том, что машина сможет однозначно отличить электронную визитную карточку партнёра по бизнесу от порнографической картинке или, сгруппировать информацию из нескольких источников на одной странице по тем или иным принципам (яркий пример этого – технология RSS).

Какой будет семантическая сеть, такой, как видят её сейчас, имеющей несколько полярных концепций, или совсем другой – не известно, но, она будет уже совсем скоро.

Технологии XML, идея описания данных метаданными и идея построения связей, собственно сети, будут воплощены в семантической сети.

То есть, умение применять логику, и навыки XML программирования – это именно то, чему надо обучать, а главное, нет никакой проблемы с тем, как это делать, более того, уже есть опыт.

Для студентов факультета информатики уральского государственного педагогического университета в рамках специализации «WEB технологии» был разработан и успешно внедрён курс «Современные XML-технологии». Поскольку программа специализации подразумевает непрерывное обучение, данный курс опирается на знания, полученные студентами при изучении курсов, ориентированных на изучение языков разметки HTML и XHTML.

Курс проектировался как максимально общий. Это делалось для достижения максимальной гибкости и прикладной ценности. После минимальных переработок, курс, отвязанный от какого-либо конкретного инструмента, может быть оформлен для обучения в среде Eclipse с последующей интеграцией новых знаний в разработку приложений на Java. Аналогично, комбинируя основные и дополнительные теоретические части курса можно максимально гибко и эффективно вписаться в любой учебный план, как с точки зрения учебного времени, так и с точки зрения отсутствия ненужных повторений уже изученного, в рамках других курсов, материала.

Литература

1. McDermott D. Critique of Pure Reason / Drew McDermott // Computational Intelligence. - 1987. - Vol. 3, Issue 1. - P. 151-160.
2. Gardner M. 1996. The universe in a handkerchief: Lewis Carroll's mathematical recreations, games, puzzles, and word plays / Martin Gardner. New York : Copernicus, 1996. - 158 p.
3. Weiser M. Some Computer Science Issues in Ubiquitous Computing / Mark Weiser // CACM. - 1993. - Vol. 36, Issue 7. - P. 74-83.
4. Williams T. An Interview with Chris Date [Электронный ресурс] / Tony Williams // O'Reilly Network. - 2005. - Режим доступа: <http://www.oreillynet.com/lpt/a/6060>